Chilean Journal of Statistics
Volume 14, Number 1 Junio 2023, Pages 65–81
DOI: 10.32372/chjs.14-01-05

Distribution Theory
Research Paper

The exponentiated power Ishita distribution: Properties, simulations, regression, and applications

Alexsandro A. Ferreira[1,*] and Gauss M. Cordeiro[1]

[1]Department of Statistics, Federal University of Pernambuco, Recife, Brazil.
(Received: 10 October 2022 · Accepted in final form: 19 May 2023)

Abstract

The new three-parameter exponentiated power Ishita distribution is introduced, and some of its mathematical properties are addressed. Its parameters are estimated by maximum likelihood, and a simulation study examines the accuracy of the estimates. A regression model is constructed based on the logarithm of the proposed distribution. The usefulness of the proposed models is proved by means of two real data sets.

Keywords: Acception-rejection method · COVID-19 · Exponentiated-G class · Ishita distribution · Maximum likelihood · Moments.

Mathematics Subject Classification: Primary 60E05 · Secondary 62F10 · 62P10.

1. Introduction

The development of new distributions is crucial in the field of statistics. By creating and refining these distributions, researchers can more accurately model and understand complex data sets. Recently, there has been a growing interest in the practice of adding parameters to known distributions. This allows for the creation of diverse shapes of hazard rate functions (HRFs), which can be used to analyze different types of data. Some distributions generated using this technique include the transmuted Dagum (Elbatal et al., 2015), Harris extended Lindley (Cordeiro et al., 2019), Mc-Donald Chen (Reis et al., 2022), and Topp-Leone log-normal (Chesneau et al., 2022) distributions.

One popular approach is to raise a cumulative distribution to an additional power parameter. This technique is commonly referred to as the exponentiated-G (exp-G) class. This class is highly flexible and versatile and has gained significant attention in recent years. Tahir and Nadarajah (2015) provided a summary of the properties of various distributions in the exp-G class in Table 1, thus including well-known distributions such as the exp-Weibull (EW) (Mudholkar and Srivastava, 1993), exp-exponential (Gupta and Kundu, 2001), exp-Fréchet (EF) (Nadarajah and Kotz, 2003), and exp-gamma (Nadarajah and Gupta, 2007). One of the main advantages of this class is that its mathematical properties can be used to obtain the properties of other classes of distributions, such as the beta-G (Eugene et al., 2002), gamma-G (Zografos and Balakrishnan, 2009), Kumaraswamy-G (Cordeiro and de Castro, 2011), and Weibull-G (Bourguignon et al., 2014) classes, among many others.

*Corresponding author. Email: alexsandro.ferreira.aaf@gmail.com.

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

Due to the limitations of the exponential and Lindley (Lindley, 1958) distributions for modeling lifetime data from biomedical science and engineering, Shanker and Shukla (2017a) introduced the Ishita (IS) distribution, based on a mixture of exponential and gamma distributions, which is capable of modeling this data set more accurately. Later, Shukla and Shanker (2018) extended this distribution using the power transformation technique, called the power Ishita (PI) distribution, with applications to real engineering data. In comparison to other distributions, such as the power Akash (Shanker and Shukla, 2017b), power Lindley (Ghitany et al., 2013), and exponential, the PI distribution has proved superior performance in terms of both accuracy and flexibility.

This paper proposes a flexible extension of the PI distribution that can be applied in several fields. The introduction of an extra shape parameter can provide better fits to various types of data. This extension represents a valuable contribution to statistical modeling, since it increases the flexibility and applicability of the PI distribution.

The remainder of the article is structured as follows: Section 2 defines the exponentiated power Ishita (EPI) distribution, and Section 3 provides some of its structural properties. The estimation of the parameters by maximum likelihood is addressed in Section 4, and a regression model is constructed in Section 5. A simulation study is done in Section 6 to examine the precision of the estimators. Two real data sets are analyzed in Section 7, and some conclusions are reported in Section 8.

2. Background

The cumulative distribution function (CDF) of the PI distribution can be expressed as (for $x > 0$)

$$G(x;\theta,\alpha) = 1 - \left[1 + \frac{\theta x^{\alpha}\,(\theta x^{\alpha}+2)}{\theta^3+2}\right] \mathrm{e}^{-\theta x^{\alpha}}\,, \tag{2.1}$$

where $\theta, \alpha > 0$. The probability density function (PDF) corresponding to Equation (2.1) is

$$g(x;\theta,\alpha) = \frac{\alpha\,\theta^3}{\theta^3+2}\left[\theta + x^{2\alpha}\right] x^{\alpha-1}\,\mathrm{e}^{-\theta x^{\alpha}}\,. \tag{2.2}$$

The PI distribution is a two-component mixture that combines a Weibull distribution and a generalized gamma distribution, with a mixing proportion $\theta^3/(\theta^3+2)$.

The CDF and PDF of the exp-G distribution with power parameter $c > 0$ are given by

$$F(x;c,\boldsymbol{\xi}) = G(x;\boldsymbol{\xi})^c\,, \tag{2.3}$$

and

$$f(x;c,\boldsymbol{\xi}) = c\,g(x;\boldsymbol{\xi})\,G(x;\boldsymbol{\xi})^{c-1}, \tag{2.4}$$

respectively, where $\boldsymbol{\xi}$ is the parameter vector of $G(\cdot)$.

By substituting Equation (2.1) in Equation (2.3) and Equation (2.1) and Equation (2.2) in Equation (2.4), the CDF and PDF of the random variable $X \sim \mathrm{EPI}\,(c,\theta,\alpha)$ having the new three-parameter EPI distribution follow as

$$F(x;c,\theta,\alpha) = \left\{1 - \left[1 + \frac{\theta x^{\alpha}\,(\theta x^{\alpha}+2)}{\theta^3+2}\right] \mathrm{e}^{-\theta x^{\alpha}}\right\}^{c}\,, \quad x > 0, \tag{2.5}$$

and

$$f(x;c,\theta,\alpha)=\frac{c\,\alpha\,\theta^3}{\theta^3+2}\left[\theta+x^{2\alpha}\right]x^{\alpha-1}\,\mathrm{e}^{-\theta x^{\alpha}}\left\{1-\left[1+\frac{\theta x^{\alpha}\,(\theta x^{\alpha}+2)}{\theta^3+2}\right]\mathrm{e}^{-\theta x^{\alpha}}\right\}^{c-1}, \quad (2.6)$$

respectively. Its HRF takes the form

$$\tau(x;c,\theta,\alpha)=\frac{\frac{c\,\alpha\,\theta^3}{\theta^3+2}\left[\theta+x^{2\alpha}\right]x^{\alpha-1}\,\mathrm{e}^{-\theta x^{\alpha}}\left\{1-\left[1+\frac{\theta x^{\alpha}(\theta x^{\alpha}+2)}{\theta^3+2}\right]\mathrm{e}^{-\theta x^{\alpha}}\right\}^{c-1}}{1-\left\{1-\left[1+\frac{\theta x^{\alpha}(\theta x^{\alpha}+2)}{\theta^3+2}\right]\mathrm{e}^{-\theta x^{\alpha}}\right\}^{c}}.$$

The exponentiated Ishita (EI) (Rather and Subramanian, 2019), PI and IS distributions are special cases of the EPI distribution when $\alpha=1$, $c=1$ and $\alpha=c=1$, respectively. Figure 1 displays plots of the PDF of X to show some shapes, including decreasing, symmetric, right-skewed, and left-skewed. Plots of the HRF of X are reported in Figure 2, which has four shapes: increasing, decreasing, unimodal, and bathtub. All plots in Sections 2 and 3 are obtained in `R` (version 4.2.3) (R Core Team, 2022). This version is utilized for all of the analyses conducted in this article.

The EPI distribution offers several motivations. This extension adds greater flexibility to the PDF and HRF, with its HRF exhibiting both bathtub and unimodal shapes. Additionally, its mathematical properties, including ordinary and incomplete moments, can be leveraged in future expansions of the PI distribution through the use of other generators such as the Beta-G, Weibull-G, Gamma-G, and Kumaraswamy-G. This is possible by the ability to determine the mathematical properties of these generators based on the properties of the exp-G class. Finally, it provides a consistently better fit to some real data sets compared to its PI, EI, and IS sub-models, as illustrated in Section 7.

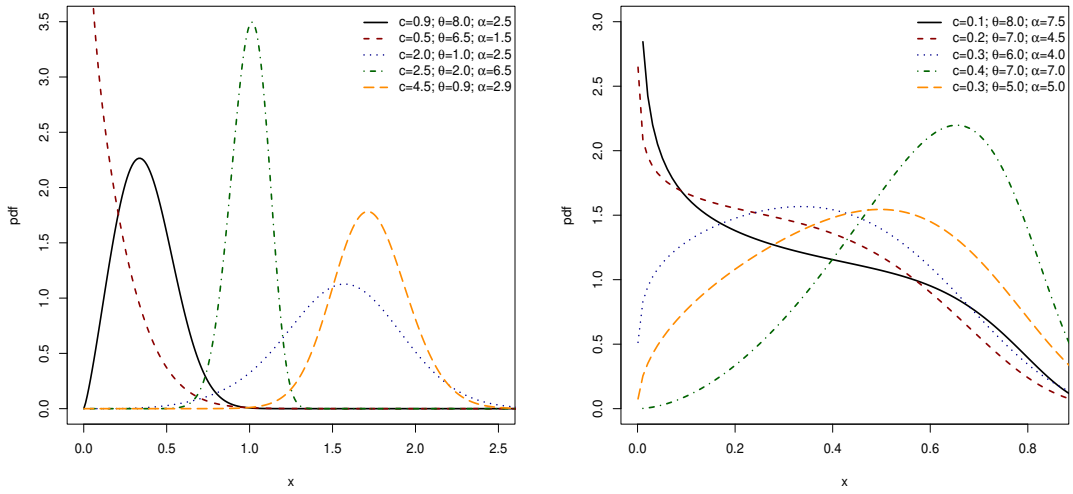


Figure 1. PDF of $X\sim\mathrm{EPI}\,(c,\theta,\alpha)$.

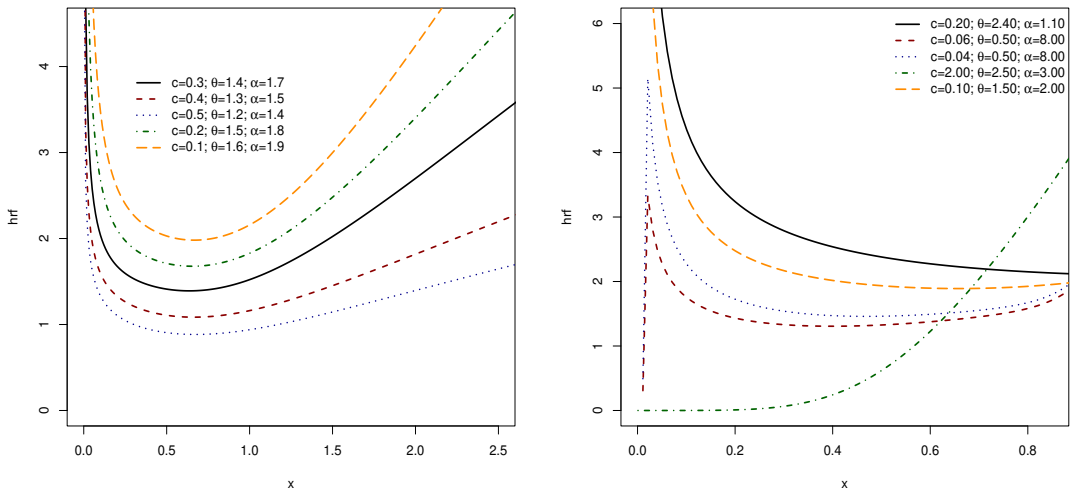


Figure 2. HRF of $X \sim \text{EPI}(c, \theta, \alpha)$.

3. Properties

The sth moment of X is defined as $\mu'_s = \mathbb{E}(X^s) = \int_0^\infty x^s f(x)\,\mathrm{d}x$. We can write from Equation (2.6)

$$\mu'_s = \frac{c\,\alpha\,\theta^3}{\theta^3+2}\int_0^\infty \left(\theta + x^{2\alpha}\right) x^{s+\alpha-1}\,\mathrm{e}^{-\theta x^\alpha}\left\{1-\left[1+\frac{\theta x^\alpha\,(\theta x^\alpha+2)}{\theta^3+2}\right]\mathrm{e}^{-\theta x^\alpha}\right\}^{c-1}\mathrm{d}x\,. \quad (3.7)$$

The integral in Equation (3.7) can be solved numerically using the programs `R`, `Mathematica`, `Maple`, or `Ox`, which have subroutines with great precision.

An analytical expression for the sth moment of X can be derived. By using the generalized Binomial Theorem on $\{1-[1+(\theta x^\alpha\,(\theta x^\alpha+2))/(\theta^3+2)]\,\mathrm{e}^{-\theta x^\alpha}\}^{c-1}$, we have

$$\mu'_s = \frac{c\,\alpha\,\theta^3}{\theta^3+2}\sum_{i=0}^{\infty}(-1)^i\binom{c-1}{i}\int_0^\infty \left(\theta + x^{2\alpha}\right) x^{s+\alpha-1}\,\mathrm{e}^{-(i+1)\theta x^\alpha}\left[1+\frac{\theta x^\alpha\,(\theta x^\alpha+2)}{\theta^3+2}\right]^i\mathrm{d}x\,.$$

Applying the Binomial Theorem on $[1+(\theta x^\alpha\,(\theta x^\alpha+2))/(\theta^3+2)]^i$ gives

$$\mu'_s = c\,\alpha\sum_{i=0}^{\infty}\sum_{j=0}^{i}(-1)^i\binom{c-1}{i}\binom{i}{j}\frac{\theta^{3+j}}{(\theta^3+2)^{j+1}}\int_0^\infty \left(\theta + x^{2\alpha}\right)(\theta x^\alpha+2)^j \times x^{\alpha(1+j)+s-1}\,\mathrm{e}^{-(i+1)\theta x^\alpha}\,\mathrm{d}x\,.$$

Again, using the Binomial Theorem on $(\theta x^\alpha+2)^j$,

$$\mu'_s = c\,\alpha\sum_{i=0}^{\infty}\sum_{j=0}^{i}\sum_{k=0}^{j}(-1)^i\,2^{j-k}\binom{c-1}{i}\binom{i}{j}\binom{j}{k}\frac{\theta^{3+j+k}}{(\theta^3+2)^{j+1}}\int_0^\infty \left(\theta + x^{2\alpha}\right) \times x^{\alpha(1+j+k)+s-1}\,\mathrm{e}^{-(i+1)\theta x^\alpha}\,\mathrm{d}x\,.$$

Replacing $\sum_{i=0}^{\infty}\sum_{j=0}^{i}$ by $\sum_{j=0}^{\infty}\sum_{i=j}^{\infty}$ gives

$$\mu'_s = c\sum_{j=0}^{\infty}\sum_{i=j}^{\infty}\sum_{k=0}^{j}(-1)^i\,2^{j-k}\binom{c-1}{i}\binom{i}{j}\binom{j}{k}\frac{\theta^{3+j+k}}{(\theta^3+2)^{j+1}} \times\left\{\frac{\theta\,\Gamma\left[1+j+k+s/\alpha\right]}{\left[(i+1)\,\theta\right]^{1+j+k+s/\alpha}}+\frac{\Gamma\left[3+j+k+s/\alpha\right]}{\left[(i+1)\,\theta\right]^{3+j+k+s/\alpha}}\right\}, \tag{3.8}$$

where $\Gamma(\cdot)$ is the gamma function.

Table 1. Numerical moments of the EPI distribution.

Equation (3.7)			
μ'_s	$c=0.5,\ \alpha=0.5$	$c=0.5,\ \alpha=2.5$	$c=0.5,\ \alpha=3.5$
μ'_1	0.2710083	0.4887077	0.5728775
μ'_2	0.8083474	0.3347754	0.4035652
μ'_3	6.0173901	0.2772582	0.3227265
μ'_4	79.2698702	0.2612512	0.2823020
μ'_s	$c=1.5,\ \alpha=0.5$	$c=1.5,\ \alpha=2.5$	$c=1.5,\ \alpha=3.5$
μ'_1	0.7037671	0.7610903	0.8113205
μ'_2	2.3492264	0.6555472	0.7044505
μ'_3	17.9042204	0.6224355	0.6471525
μ'_4	237.2658052	0.6400585	0.6237897
Equation (3.8)			
μ'_s	$c=0.5,\ \alpha=0.5$	$c=0.5,\ \alpha=2.5$	$c=0.5,\ \alpha=3.5$
μ'_1	0.2709654	0.4722906	0.5441371
μ'_2	0.8083471	0.3319362	0.3960836
μ'_3	6.0173898	0.2766467	0.3204721
μ'_4	79.2698679	0.2610981	0.2815484
μ'_s	$c=1.5,\ \alpha=0.5$	$c=1.5,\ \alpha=2.5$	$c=1.5,\ \alpha=3.5$
μ'_1	0.7037700	0.7618455	0.8125474
μ'_2	2.3492264	0.6557035	0.7048268
μ'_3	17.9042204	0.6224730	0.6472789
μ'_4	237.2658052	0.6400686	0.6238354

The first four moments of X from Equations (3.7) and (3.8) for some values of c and α with $\theta = 2.5$ are listed in Table 1. These moments are quite close, thus demonstrating the accuracy of Equation (3.8) with just the first 15 terms. Equations (3.7) and (3.8) are computed using `R` and `Mathematica` (version 12.2) (Wolfram Research, 2020) programs.

The skewness and kurtosis of X are easily determined from the first four ordinary moments. Plots of these quantities as functions of c (for some values of θ and α) are reported in Figure 3, which can be increasing, decreasing, decreasing-increasing, and increasing-decreasing.

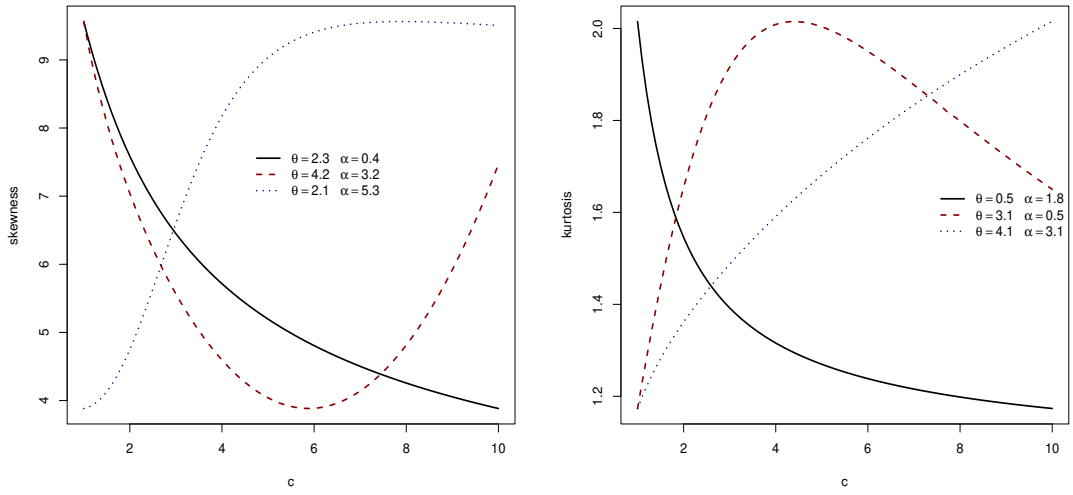


Figure 3. Skewness and kurtosis of $X \sim \mathrm{EPI}(c, \theta, \alpha)$.

Further, the sth incomplete moment of X, say $m_s(z) = \int_0^z x^s f(x)\,\mathrm{d}x$, can be expressed as

$$m_s(z) = \frac{c\,\alpha\,\theta^3}{\theta^3+2}\int_0^z \left(\theta + x^{2\alpha}\right) x^{s+\alpha-1}\,\mathrm{e}^{-\theta x^\alpha}\left\{1-\left[1+\frac{\theta x^\alpha\,(\theta x^\alpha+2)}{\theta^3+2}\right]\mathrm{e}^{-\theta x^\alpha}\right\}^{c-1}\mathrm{d}x\,.$$

Following the same steps that lead to Equation (3.8), we have

$$m_s(z) = c\sum_{j=0}^{\infty}\sum_{i=j}^{\infty}\sum_{k=0}^{j}(-1)^i\,2^{j-k}\binom{c-1}{i}\binom{i}{j}\binom{j}{k}\frac{\theta^{3+j+k}}{(\theta^3+2)^{j+1}}$$
$$\times\left\{\frac{\theta\,\gamma\left[\frac{\alpha(1+j+k)+s}{\alpha},(i+1)\theta z^\alpha\right]}{[(i+1)\,\theta]^{\frac{\alpha(1+j+k)+s}{\alpha}}}+\frac{\gamma\left[\frac{\alpha(3+j+k)+s}{\alpha},(i+1)\theta z^\alpha\right]}{[(i+1)\,\theta]^{\frac{\alpha(3+j+k)+s}{\alpha}}}\right\},$$

where $\gamma(\cdot,\cdot)$ denotes the lower incomplete gamma function.

The first incomplete moment $m_1(z)$ can be used to calculate the total deviations from the mean and median of X. These can be expressed as $\delta_1 = 2\mu_1' F(\mu_1') - 2m_1(\mu_1')$ and $\delta_2 = \mu_1' - 2m_1(\mathrm{M})$, respectively, where M can be found from the CDF given in Equation (2.5) by setting $F(\mathrm{M}) = 0.5$.

For a given probability ν, the Bonferroni and Lorenz curves of X are $B(\nu) = m_1(q)/\nu\mu_1'$ and $L(\nu) = m_1(q)/\mu_1'$, respectively, where q is the solution of $F(q) = \nu$. Plots of these curves for X versus ν for selected values of c and θ ($\alpha = 1.5$) are reported in Figure 4.

4. Estimation

Let $x_1, x_2, \ldots, x_n$ be independent and identically distributed (IID) observations from the EPI distribution, and let $\boldsymbol{\zeta} = (c, \theta, \alpha)^\top$ be the vector of unknown parameters. The log-

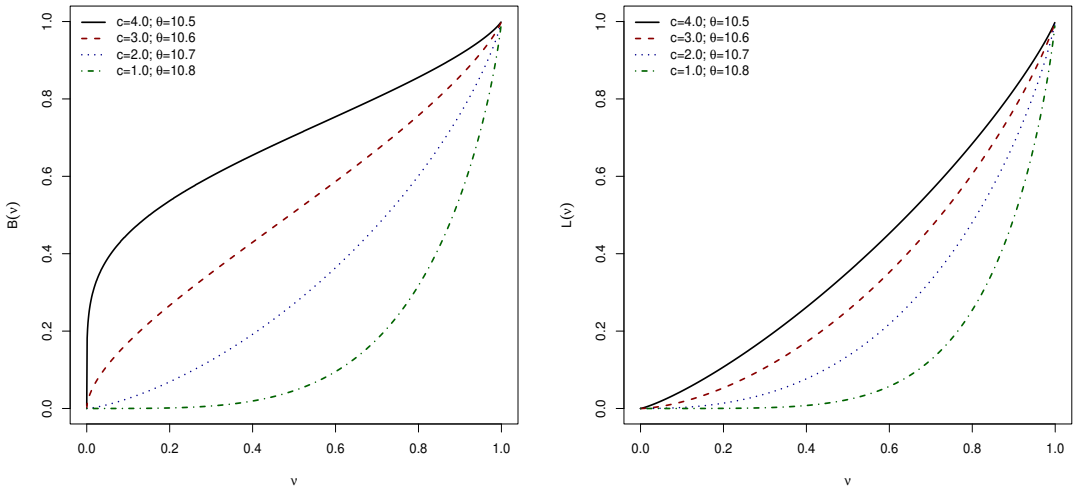


Figure 4. Bonferroni and Lorenz curves of $X \sim \text{EPI}(c, \theta, \alpha)$.

likelihood function for $\boldsymbol{\zeta}$ is

$$\ell(\boldsymbol{\zeta}) = n\left[\log(c) + \log(\alpha) + 3\log(\theta) - \log(\theta^3+2)\right] + \sum_{i=1}^{n}\log(\theta + x_i^{2\alpha}) + (\alpha-1)\sum_{i=1}^{n}\log(x_i)$$
$$- \theta\sum_{i=1}^{n} x_i^{\alpha} + (c-1)\sum_{i=1}^{n}\log\left\{1 - \left[1 + \frac{\theta x_i^{\alpha}\,(\theta x_i^{\alpha}+2)}{\theta^3+2}\right]\mathrm{e}^{-\theta x_i^{\alpha}}\right\}. \tag{4.9}$$

The maximum likelihood estimate (MLE) $\hat{\boldsymbol{\zeta}}$ of $\boldsymbol{\zeta}$ is determined by maximizing Equation (4.9) numerically using statistical programs like `R` (`optim` function), `Ox` (`MaxBFGS` function), and `SAS` (`PROC NLMIXED` function).

5. Regression

By applying the transformation $Y = \log(X)$, where X has PDF given in Equation (2.6) and setting $\alpha = 1/\sigma$ and $\theta = \mathrm{e}^{-\mu/\sigma}$, the log-exponentiated power Ishita (LEPI) density has the form (for $y \in \mathbb{R}$)

$$f(y; c, \sigma, \mu) = \frac{c\,\mathrm{e}^{\left(\frac{y-3\,\mu}{\sigma}\right)}\left(\mathrm{e}^{-\mu/\sigma} + \mathrm{e}^{2\,y/\sigma}\right)\mathrm{e}^{-\mathrm{e}^{\left(\frac{y-\mu}{\sigma}\right)}}}{\sigma\,\left(\mathrm{e}^{-3\,\mu/\sigma}+2\right)}\left\{1 - \left[1 + \frac{\mathrm{e}^{\left(\frac{y-\mu}{\sigma}\right)}\left(\mathrm{e}^{\left(\frac{y-\mu}{\sigma}\right)}+2\right)}{\left(\mathrm{e}^{-3\,\mu/\sigma}+2\right)}\right]\mathrm{e}^{-\mathrm{e}^{\left(\frac{y-\mu}{\sigma}\right)}}\right\}^{c-1}, \tag{5.10}$$

where c, $\sigma > 0$ and $\mu \in \mathbb{R}$. So, if $X \sim \text{EPI}(c, \theta, \alpha)$, then $Y = \log(X) \sim \text{LEPI}(c, \sigma, \mu)$.

The survival and density functions $Z = (Y-\mu)/\sigma$ are

$$S(z; c, \sigma, \mu) = 1 - \left\{1 - \left[1 + \frac{\mathrm{e}^{z}\,(\mathrm{e}^{z}+2)}{\left(\mathrm{e}^{-3\,\mu/\sigma}+2\right)}\right]\mathrm{e}^{-\mathrm{e}^{z}}\right\}^{c}$$

and

$$f(z; c, \sigma, \mu) = \frac{c\,w(z)\,\mathrm{e}^{-\mathrm{e}^{z}}}{\left(\mathrm{e}^{-3\,\mu/\sigma}+2\right)}\left\{1 - \left[1 + \frac{\mathrm{e}^{z}\,(\mathrm{e}^{z}+2)}{\left(\mathrm{e}^{-3\,\mu/\sigma}+2\right)}\right]\mathrm{e}^{-\mathrm{e}^{z}}\right\}^{c-1}, \quad z \in \mathbb{R}, \tag{5.11}$$

respectively, where

$$w(z) = \mathrm{e}^{\left(\frac{z\sigma-2\,\mu}{\sigma}\right)}\left(\mathrm{e}^{-\mu/\sigma} + \mathrm{e}^{2(z\sigma+\mu)/\sigma}\right).$$

Parametric regression models are often used for censored data. In this context, based on Equation (5.10), a regression model for the response variable Y_i and a vector of explanatory variables $\boldsymbol{v}_i^\top = (v_{i1}, v_{i2}, \ldots, v_{ip})$ is constructed as

$$y_i = \boldsymbol{v}_i^\top \boldsymbol{\beta} + \sigma\, z_i\,,\ i = 1, \ldots, n\,, \tag{5.12}$$

where $\mu_i = \boldsymbol{v}_i^\top \boldsymbol{\beta}$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^\top$ is a vector of unknown coefficients, and z is the random error with density given in Equation (5.11).

The density and survival functions of $Y_i|\boldsymbol{v}_i$ are

$$f(y|\boldsymbol{v}_i) = \frac{c\, w(z_i)\, \mathrm{e}^{-\mathrm{e}^{z_i}}}{\sigma\left(\mathrm{e}^{-3\,\mu_i/\sigma} + 2\right)}\left\{1 - \left[1 + \frac{\mathrm{e}^{z_i}\left(\mathrm{e}^{z_i} + 2\right)}{\left(\mathrm{e}^{-3\,\mu_i/\sigma} + 2\right)}\right]\mathrm{e}^{-\mathrm{e}^{z_i}}\right\}^{c-1} \tag{5.13}$$

and

$$S(y|\boldsymbol{v}_i) = 1 - \left\{1 - \left[1 + \frac{\mathrm{e}^{z_i}\left(\mathrm{e}^{z_i} + 2\right)}{\left(\mathrm{e}^{-3\,\mu_i/\sigma} + 2\right)}\right]\mathrm{e}^{-\mathrm{e}^{z_i}}\right\}^{c},$$

respectively, where $w(z_i) = \mathrm{e}^{\left(\frac{z_i\sigma-2\,\mu_i}{\sigma}\right)}\left(\mathrm{e}^{-\mu_i/\sigma} + \mathrm{e}^{2(z_i\sigma+\mu_i)/\sigma}\right)$ and $z_i = (y_i - \mu_i)/\sigma$.

Maximum likelihood is adopted to estimate the parameters in Equation (5.12) for right-censored data. Let Y_i and C_i be the lifetime and the non-informative censoring time (assuming independence), respectively, and $y_i = \min(Y_i, C_i)$. Then, the log-likelihood function for $\boldsymbol{\eta} = (c, \sigma, \boldsymbol{\beta}^\top)^\top$ is

$$\begin{aligned}\ell(\boldsymbol{\eta}) = {} & d\,[\log(c) - \log(\sigma)\,] - \sum_{i\in F}\log\left(\mathrm{e}^{-3\mu_i/\sigma} + 2\right) + \sum_{i\in F}\log[w(z_i)] - \sum_{i\in F}\mathrm{e}^{z_i} \\ & + (c-1)\sum_{i\in F}\log\left\{1 - \left[1 + \frac{\mathrm{e}^{z_i}\left(\mathrm{e}^{z_i} + 2\right)}{\left(\mathrm{e}^{-3\,\mu_i/\sigma} + 2\right)}\right]\mathrm{e}^{-\mathrm{e}^{z_i}}\right\} \\ & + \sum_{i\in C}\log\left[1 - \left\{1 - \left[1 + \frac{\mathrm{e}^{z_i}\left(\mathrm{e}^{z_i} + 2\right)}{\left(\mathrm{e}^{-3\,\mu_i/\sigma} + 2\right)}\right]\mathrm{e}^{-\mathrm{e}^{z_i}}\right\}^{c}\right],\end{aligned} \tag{5.14}$$

where F and C are the sets of uncensored and censored observations, respectively, and d is the number of failures. The MLE $\hat{\boldsymbol{\eta}}$ of the vector of unknown parameters can be found by maximizing Equation (5.14).

6. Simulations

The EPI distribution is simulated under three different scenarios to examine the accuracy of the MLEs. The acceptance-rejection method is adopted to generate random samples of sizes $n = 50, 100, 300$, and 600 from this distribution. The process is repeated 1,000 times and the average estimates (AEs), biases, and mean squared errors (MSEs) are computed.

The algorithm for generating random samples uses the acceptance-rejection method:

(1) Generate t from the density $h(t) = \alpha\,\theta\, t^{\alpha-1}\mathrm{e}^{-\theta\, t^{\alpha}}$.
(2) Generate $u \sim$ uniform $(0, 1)$.
(3) If $u \leq f(t)/Nh(t)$, set $x = t$, where $f(\cdot)$ is the PDF given in Equation (2.6) and $N = \max\,[f(t)/h(t)\,]$. Otherwise, return to step 1.

Figure 5 displays graphically the approximation of the acceptance-rejection method. The estimated PDF and CDF of the EPI distribution are very close to the histogram and empirical CDF of the generated samples, thus indicating a good performance of this method.

Table 2. Simulation findings from the EPI distribution.

		(2.0, 10, 0.5)			(1.4, 7.0, 5.0)			(3.0, 9.0, 3.5)		
n	$\boldsymbol{\zeta}$	AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
50	c	3.3155	1.3155	15.8513	2.1312	0.7312	8.5224	4.7897	1.7897	39.0636
	θ	13.0738	3.0738	66.8723	8.5586	1.5586	14.3504	11.4107	2.4107	32.4455
	α	0.5598	0.0598	0.0582	5.7608	0.7608	6.8695	4.1010	0.6010	3.4317
100	c	2.6372	0.6372	5.4844	1.6828	0.2828	1.4734	4.2802	1.2802	20.6862
	θ	11.2604	1.2604	16.9159	7.7355	0.7355	5.7606	9.9202	0.9202	6.0715
	α	0.5250	0.0250	0.0245	5.3505	0.3505	2.5289	3.7087	0.2087	1.4281
300	c	2.1606	0.1606	0.6441	1.4539	0.0539	0.1919	3.2622	0.2622	2.0005
	θ	10.3520	0.3520	2.1693	7.1913	0.1913	0.4788	9.2594	0.2594	0.8096
	α	0.5085	0.0085	0.0069	5.1220	0.1220	0.5843	3.5798	0.0798	0.3345
600	c	2.0609	0.0609	0.2186	1.4374	0.0374	0.0969	3.1098	0.1098	0.7249
	θ	10.1348	0.1348	0.7482	7.1046	0.1046	0.2178	9.1221	0.1221	0.3645
	α	0.5037	0.0037	0.0029	5.0560	0.0560	0.3241	3.5399	0.0399	0.1625

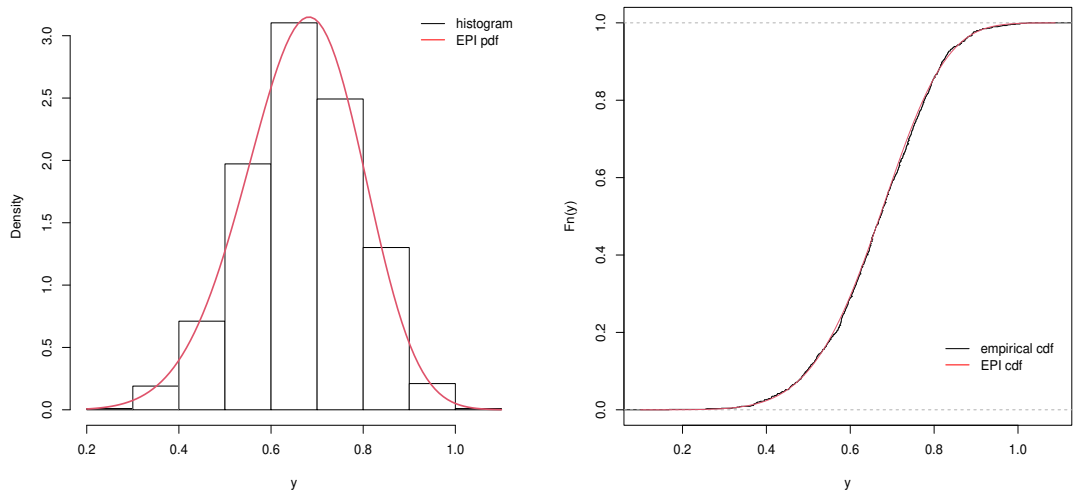


Figure 5. Histogram and empirical CDF of the generated samples. Estimated PDF and CDF for the scenario (1.4, 7.0, 5.0).

The numbers in Table 2 indicate that the AEs converge to the true parameters and that the biases and MSEs tend to zero when n increases, which proves the consistency of the EPI estimators. Note that for the $(2.0, 10, 0.5)$ scenario the parameter estimates are less accurate (except for α), while for the $(1.4, 7.0, 5.0)$ scenario, the parameter estimates are more accurate (for c and θ). Overall, the simulation results suggest that larger sample sizes and the appropriate choice of $\boldsymbol{\zeta}$ are crucial for accurate parameter estimation of the EPI distribution.

The accuracy of the MLEs in the new regression model is investigated from 1,000 samples for $c = 0.5, \sigma = 0.8, \beta_0 = 0.3$, $\beta_1 = 0.5$, and $n = 50, 100, 300$, and 600. The observations are generated from $Y_i \sim \text{LEPI}(c, \sigma, \mu_i)$, where $\mu_i = \beta_0 + \beta_1 v_{i1}$ and $v_{i1} \sim \text{uniform}(0, 1)$. The algorithm for generating random samples from the LEPI regression is similar to the EPI distribution. First, generate $v_{i1} \sim \text{uniform}(0, 1)$ and calculate $\mu_i = \beta_0 + \beta_1 v_{i1}$. Then, write $h(t_i) = (1/\sigma)\exp[(t_i - \mu_i)/\sigma]\exp\{-\exp[(t_i - \mu_i)/\sigma]\}$ and $f(\cdot)$ is the PDF given in Equation (5.13).

The findings in Table 3 show that the AEs converge to the true parameters and the biases and MSEs decay to zero when n increases (in all parameters). This fact proves that the LEPI regression estimators are consistent. In general, the estimates for β_0 and β_1 have the highest biases

and MSE values, while the estimates of c and σ have the lowest values of these measures. These simulations are done using a script with the `optim` subroutine in `R` program. All computations, including those presented in Section 7, are conducted on a Windows 10 computer with an i7-7700k processor operating at 4.20 GHz, and equipped with 16GB of memory.

Table 3. Simulation findings from the LEPI regression.

n	Parameter	AE	Bias	MSE
50	c	0.6618	0.1618	0.2711
	σ	0.7594	-0.0405	0.0452
	β_0	0.2378	-0.0621	0.2695
	β_1	0.4898	-0.0101	0.5688
100	c	0.5614	0.0614	0.0923
	σ	0.7547	-0.0452	0.0333
	β_0	0.2782	-0.0217	0.2136
	β_1	0.5436	0.0436	0.2403
300	c	0.5172	0.0172	0.0245
	σ	0.7743	-0.0256	0.0119
	β_0	0.2837	-0.0162	0.0538
	β_1	0.5445	0.0445	0.0712
600	c	0.5138	0.0138	0.0124
	σ	0.7854	-0.0145	0.0059
	β_0	0.2852	-0.0147	0.0307
	β_1	0.5268	0.0268	0.0325

7. Applications

The suitability of the proposed models is empirically proved by means of two real data applications. Model selection criteria are based on the Cramér-von Mises (W^*) and Anderson-Darling (A^*) statistics defined by Chen and Balakrishnan (1995). In addition, we consider the Akaike information criterion (AIC), Consistent Akaike information criterion (CAIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion (HQIC), and Kolmogorov-Smirnov (KS) (and its p-value). The lower the value of these statistics, the stronger the evidence of a good fit. Graphical analysis is also crucial for identifying the best fitting model. This includes analysis of data histograms, estimated PDFs and CDFs, and the empirical CDF calculated using the Kaplan-Meier method (Kaplan and Meier, 1958). The `AdequacyModel` package (Marinho et al., 2019) in `R` is used to obtain the results in Section 7.1, whereas those in Section 7.2 are found using a script developed in `R` with the `optim` function.

7.1 Cancer data

The data set comprises remission times (in months) for a random sample of 128 bladder cancer patients (Lee, 2003). The data are: 0.08, 2.09, 3.48, 4.87, 6.94 , 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50, 2.46 , 3.64, 5.09, 7.26, 9.47, 14.24, 25.82, 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.76, 26.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64, 3.88,5.32, 7.39, 10.34, 14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59, 10.66, 15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.26, 2.83, 4.33, 5.49, 7.66, 11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34, 5.71, 7.93, 11.79, 18.10, 1.46, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25, 8.37, 12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 6.76, 12.07, 21.73, 2.07, 3.36, 6.93, 8.65, 12.63, 22.69.

Some descriptive statistics for these data are reported in Table 4. The standard deviation (SD) is greater than the mean and median, and it can be deduced that the data are right-skewed and leptokurtic.

Table 4. Descriptive to cancer data.

Mean	Median	SD	Variance	Skewness	Kurtosis	Min.	Max.
9.3656	6.3950	10.4670	109.560	3.2866	18.4830	0.080	79.0501

The proposed model is compared with the PI, EF, EW, EI, and IS distributions. The EW and EF distributions are well-known nowadays. Table 5 lists the MLEs and their standard errors (SEs) from the fitted models to the current data. The results indicate that all models provide accurate estimates (except for EF).

Table 5. Findings from the fitted models to cancer data.

Model	MLEs (SEs)		
EPI (c, θ, α)	3.8899 (1.9327)	1.3690 (0.2862)	0.4955 (0.0628)
EW (c, β, α)	3.4207 (1.5491)	0.6583 (0.1116)	2.7156 (0.9872)
EF (c, β, α)	18.0663 (6.4246)	0.3296 (0.0298)	210.2781 (131.0810)
PI (θ, α)	0.6448 (0.0494)	0.6905 (0.0326)	
EI (c, θ)	0.3915 (0.0479)	0.1897 (0.0199)	
IS (θ)	0.3209 (0.0160)		

According to the results in Table 6, which includes only distributions with accurate estimates, the EPI model has the lowest values of the adequacy measures, particularly W^*, A^*, and KS, thus indicating that it provides the best fit to the COVID-19 data among the fitted models. The EW model ranks second in terms of these measures. In contrast, the EI and IS models have the highest values for W^* and A^*, as well as for other measures, suggesting a comparatively poorer fit to the data.

Since the PI, EI, and IS distributions are special cases of the EPI distribution, the likelihood ratio (LR) tests are presented in Table 7. In all three cases, the null hypotheses are rejected with low p-values, thus indicating that the EPI distribution provides a significantly better fit to the data than any of the other three distributions. This suggests that the EPI distribution is a more appropriate model for this data set.

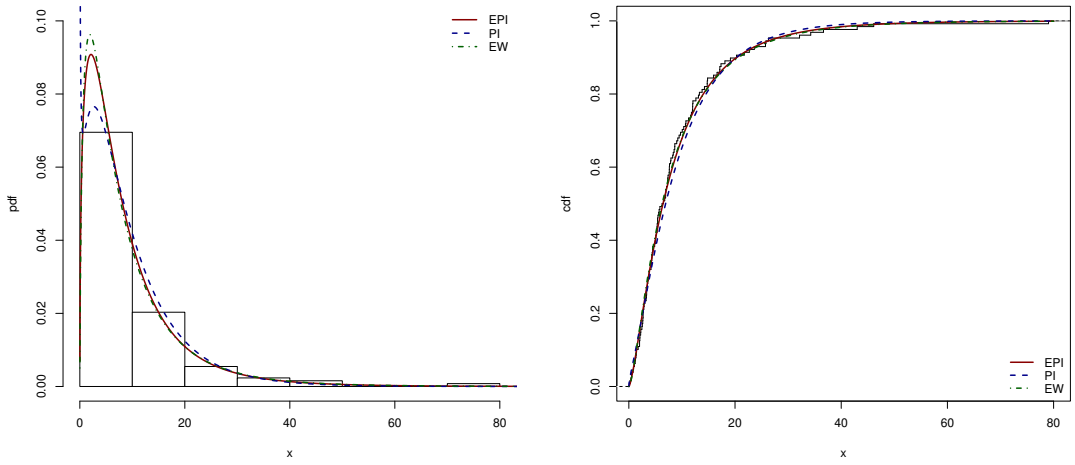
Figure 6 illustrates that the PDF and CDF of the EPI distribution are closer to the histogram and empirical CDF of the data, and Figure 7 displays the quantile-quantile (QQ) plots, thus revealing that the points are closest to the diagonal line for the EPI distribution. Based on these results, the EPI distribution can be selected as the best model.

Table 6. Some measures for the fitted models to cancer data.

Model	W^*	A^*	AIC	CAIC	BIC	HQIC	KS	p-value
EPI	0.039	0.255	826.673	826.867	835.229	830.150	0.043	0.970
EW	0.044	0.290	827.381	827.574	835.937	830.857	0.048	0.923
PI	0.115	0.692	830.477	830.573	836.181	832.795	0.071	0.529
EI	0.199	1.216	839.661	839.757	845.365	841.978	0.101	0.141
IS	0.160	0.984	903.367	903.399	906.219	904.526	0.200	< 0.001

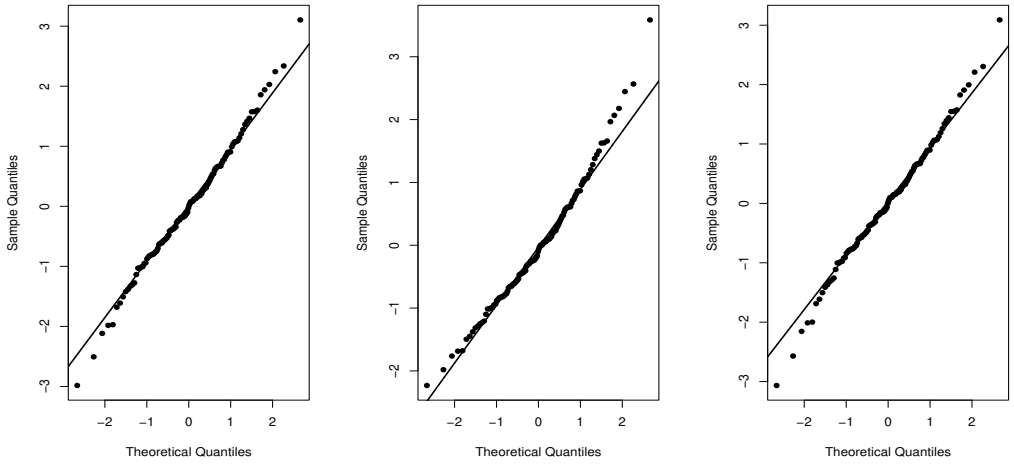
Table 7. LR tests for cancer data.

Model	Hypotheses	LR	p-value
EPI vs PI	$H_0 : c = 1$ vs $H_1 : H_0$ is false	5.8041	0.0159
EPI vs EI	$H_0 : \alpha = 1$ vs $H_1 : H_0$ is false	14.9876	0.0001
EPI vs IS	$H_0 : c = \alpha = 1$ vs $H_1 : H_0$ is false	80.6940	< 0.0001



(a) (b)

Figure 6. Estimated PDFs (a) and CDFs (b) for Cancer data.



(a) (b) (c)

Figure 7. EPI QQ-Plot (a); PI QQ-Plot (b); and EW QQ-Plot (c) for Cancer data.

7.2 COVID-19 DATA

This dataset comprises the lifetime (in days) of 322 individuals who were diagnosed with COVID-19 via RT-PCR screening in Campinas, Brazil. These data were previously analyzed by Rodrigues et al. (2022) and are available at https://github.com/gabrielamrodrigues/OLLW/blob/main/data.covid.txt. The response variable y_i represents the time elapsed from the onset of symptoms until death due to COVID-19 (failure).

In this sample, around 66.45% of the observations are censored. The variables considered (for $i = 1, \ldots, 322$) include: δ_i : censoring indicator (0 = censored, 1 = observed lifetime), v_{i1} : age (in years), and v_{i2} : diabetes mellitus (1 = yes, 0 = no or not informed). Figure 8(a) shows that individuals between the ages of 60 and 90 have the highest frequency of hospitalizations, while Figure 8(b) reveals that individuals with diabetes have a higher risk of death. The proposed regression model for these data is written as

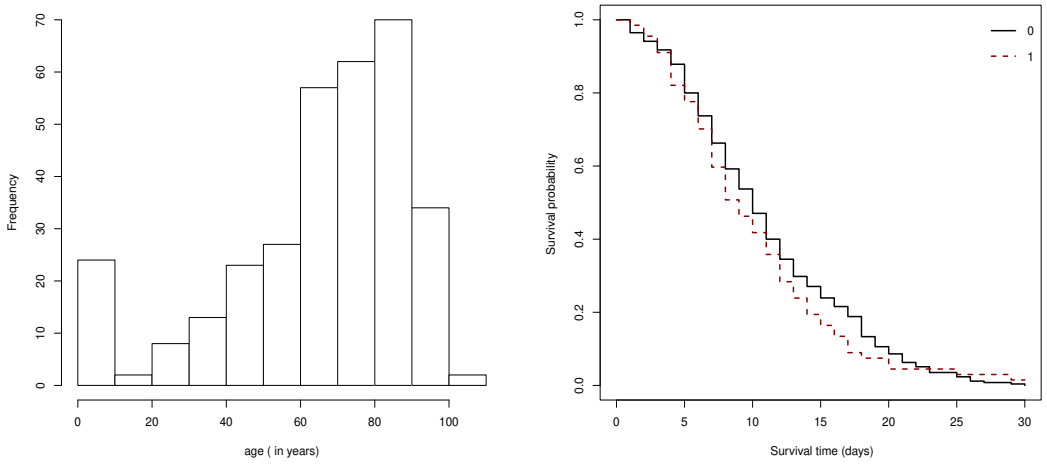
$$y_i = \beta_0 + \beta_1 v_{i1} + \beta_2 v_{i2} + \sigma z_i, \quad i = 1, \ldots, 322\,,$$

where z_i follows the pdf given in Equation (5.11). The results are compared with the log-exponentiated Weibull (LEW) (Hashimoto et al., 2010), log-exponentiated Frechét (LEF) (Al-Amoudi et al., 2016), and log-PI (LPI) regressions. The numbers in Table 8 show that the explanatory variables age and diabetes mellitus are significant at the 5% level. The negative signs of β_1 and β_2 mean that older individuals or those with diabetes tend to have shorter failure times.

According to Table 9, the LEPI regression has the lowest criterion values. The generalized likelihood ratio (GLR) test (Vuong, 1989) is used to compare the LEPI regression against the LPI (GLR = 14.1684), LEF (GLR = 10.0742), and LEW (GLR = 16.8083) regressions for a significance level of 5%. These results confirm that the LEPI model provides a better fit to the current data. For a residual analysis of this fitted regression, we adopt the quantile residuals (qrs) defined by (Dunn and Smyth, 1996)

$$qr_i = \Phi^{-1}\left\{\left(1 - \left[1 + \frac{e^{\hat{z}_i}\left(e^{\hat{z}_i} + 2\right)}{\left(e^{-3\,\hat{\mu}_i/\hat{\sigma}} + 2\right)}\right] e^{-e^{\hat{z}_i}}\right)^{\hat{c}}\right\},$$

where $\Phi(\cdot)^{-1}$ is the standard normal quantile function, $\hat{z}_i = (y_i - \hat{\mu}_i)/\hat{\sigma}$, and $\hat{\mu}_i = \boldsymbol{v}_i^{\top}\hat{\boldsymbol{\beta}}$. Only one observation is outside the range [-3,3] according to Figure 9(a), which also shows that the qrs are randomly distributed. The normal probability plot in Figure 9(b) shows that the qrs approximately follow a standard normal distribution. Therefore, there is no evidence against the LEPI regression assumptions.



(a) (b)

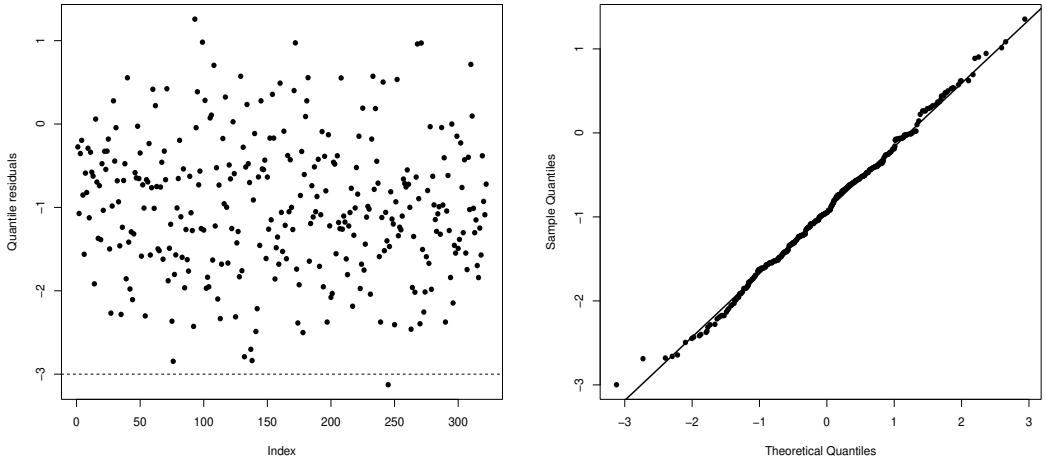
Figure 8. Histogram for age (a); and Kaplan-Meier curves for diabetes mellitus (b) for COVID-19 data.

Table 8. Findings from the fitted models to COVID-19 data.

Model	c	σ	β_0	β_1	β_2
LEPI	0.2070 (0.0887)	0.3971 (0.1460)	4.4367 (0.2365)	-0.0180 (0.0037) [< 0.0001]	-0.2439 (0.1162) [< 0.0365]
LPI	1	1.1926 (0.0774)	3.2401 (0.3020)	-0.0197 (0.0039) [< 0.0001]	-0.2848 (0.1296) [< 0.0287]
LEF	154.1795 (116.3051)	3.8146 (0.4878)	10.9723 (1.3599)	-0.0212 (0.0041) [< 0.0001]	-0.3014 (0.1453) [< 0.0389]
LEW	0.8877 (0.7995)	0.5106 (0.3514)	4.5518 (0.3585)	-0.0182 (0.0045) [0.0008]	-0.2803 (0.1265) [0.0274]

Table 9. Some measures of the fitted models to COVID-19 data.

Model	AIC	CAIC	BIC	HQIC
LEPI	428.959	429.316	447.832	436.494
LPI	433.594	433.860	448.692	439.621
LEF	441.908	442.265	460.781	449.443
LEW	429.849	430.206	448.722	437.383



(a) (b)

Figure 9. Index plot (a); and Normal probability plot (b) for Covid-19 data.

8. Conclusions, limitations, and future research

We introduced a new three-parameter exponentiated power Ishita model, which includes as special cases the power Ishita, exponentiated Ishita, and Ishita distributions. Some of its structural properties were studied. Some simulation results showed that the maximum likelihood estimators are consistent. Additionally, based on its log-transform, we constructed a regression model for censored data and demonstrated its applicability to the COVID-19 data set. Compared to other established models, our model has proven to have the best fit. The results showed that advanced age or pre-existing diabetes are significant factors affecting survival time.

A limitation of the exponentiated power Ishita distribution is the absence of an analytic solution for its quantile function. As a result, numerical methods must be utilized to estimate the quantile values. This limitation also impacts the generation of random samples from this distribution, thus resulting in a process that can be slower and less precise (depending on parameter selection) compared to distributions with analytical solutions for their quantile functions.

Future work could be directed towards building other regression models based on this distribution, drawing on the approaches outlined in recent studies by Prataviera et al. (2018), Biazatti et al. (2022), and Rodrigues et al. (2022).

Supplementary materials

The computational routine implemented in `R` is available online at https://github.com/alexaaf31/Exponentiated-power-Ishita-distribution.R.git.

Author contributions Conceptualization, A.A.F., G.M.C.; methodology, A.A.F., G.M.C.; software, A.A.F.; validation, A.A.F., G.M.C.; formal analysis, A.A.F.; investigation, A.A.F.; data curation, A.A.F.; writing-original draft preparation, A.A.F.; writing-review and editing, G.M.C.; visualization, A.A.F., G.M.C.; supervision, G.M.C. All authors have read and agreed the published version of the paper.

Acknowledgements The authors would like to express their gratitude to the Editors-in-Chief and the anonymous reviewers for their valuable comments and suggestions, which have greatly improved the quality of the paper.

Funding This work was supported by the Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) [IBPG-1448-1.02/20].

Conflicts of interest The authors declare no conflict of interest.

References

Al-Amoudi, H.H., Mousa, S.A., and Baharith, L.A., 2016. Log-exponentiated Frechét regression model with censored data. International Journal of Advances in Applied Sciences, 3, 1–9.

Biazatti, E.C., Cordeiro, G.M., Rodrigues, G.M., Ortega, E.M.M., and de Santana, L.H., 2022. A Weibull-beta prime distribution to model COVID-19 data with the presence of covariates and censored data. Stats, 5, 1159–1173.

Bourguignon, M., Silva, R.B., and Cordeiro, G.M., 2014. The Weibull-G family of probability distributions. Journal of Data Science, 12, 53–68.

Chen, G. and Balakrishnan, N., 1995. A general purpose approximate goodness-of-fit test. Journal of Quality Technology, 27, 154–161.

Chesneau, C., Irshad, M.R., Shibu, D.S., Nitin, S.L., and Maya, R., 2022. On the Topp-Leone log-normal distribution: Properties, modeling, and applications in astronomical and cancer data. Chilean Journal of Statistics, 13, 67–90.

Cordeiro, G.M. and de Castro, M., 2011. A new family of generalized distributions. Journal of Statistical Computation and Simulation, 81, 883–898.

Cordeiro, G.M., Mansoor, M., and Provost, S.B., 2019. The Harris extended Lindley distribution for modeling hydrological data. Chilean Journal of Statistics, 10, 77–94.

Dunn P.K. and Smyth G.K., 1996. Randomized quantile residuals. Journal of Computational and Graphical Statistics, 5, 236–244.

Elbatal, I. and Aryal, G., 2015. Transmuted Dagum distribution with applications. Chilean Journal of Statistics, 6, 31–45.

Eugene, N., Lee, C., and Famoye, F., 2002. Beta-normal distribution and its applications. Communications in Statistics: Theory and Methods, 31, 497–512.

Ghitany, M.E., Al-Mutairi, D.K., Balakrishnan, N., and Al-Enezi, L.J., 2013. Power Lindley distribution and associated inference. Computational Statistics and Data Analysis, 64, 20–33.

Gupta R.D., and Kundu D., 2001. Exponentiated exponential family: an alternative to gamma and Weibull distributions. Biometrical Journal, 43, 117–130.

Hashimoto, E.M., Ortega, E.M.M., Cancho, V.G., and Cordeiro, G.M., 2010. The log-exponentiated Weibull regression model for interval-censored data. Computational Statistics and Data Analysis, 54, 1017–1035.

Kaplan, E.L. and Meier, P., 1958. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53, 457–481.

Lee, E.T. and Wang, J.W., 2003. Statistical Methods for Survival Data Analysis. Wiley, New York.

Lindley, D.V., 1958. Fiducial distributions and Bayes' theorem. Journal of the Royal Statistical Society B, 20(1), 102–107.

Marinho P.R.D., Silva R.B., Bourguignon M., Cordeiro G.M., and Nadarajah S., 2019. AdequacyModel: An `R` package for probability distributions and general purpose optimization. Plos ONE, 14, e0221487.

Mudholkar G. and Srivastava D., 1993. Exponentiated Weibull family for analyzing bathtub failure-rate data. IEEE Transactions on Reliability, 42, 299–302.

Nadarajah S. and Gupta A.K., 2007. The exponentiated gamma distribution with application to drought data. Calcutta Statistical Association Bulletin, 59, 29–54.

Nadarajah S. and Kotz S., 2003. The exponentiated Fréchet distribution. Interstat Electronic Journal, 14, 01–07.

Prataviera, F., Ortega, E.M.M., Cordeiro, G.M., Pescim, R.R., and Verssani, B.A., 2018. A new generalized odd log-logistic flexible Weibull regression model with applications in repairable systems. Reliability Engineering and System Safety, 176, 13–26.

R Core Team: A Language and Environment for Statistical Computing, 2022. R Foundation for Statistical Computing, Vienna, Austria.

Rather A.A. and Subramanian C., 2019. Exponentiated Ishita distribution: Properties and applications. International Journal of Management, Technology and Engineering, 9, 2473–2484.

Ribeiro-Reis, L.D., Cordeiro, G.M., and de Santana e Silva, J.J., 2022. The Mc-Donald Chen distribution: A new bimodal distribution with properties and applications. Chilean Journal of Statistics, 13, 91–111.

Rodrigues, G.M., Ortega, E.M.M., Cordeiro, G.M., and Vila, R., 2022. An extended Weibull regression for censored data: Application for COVID-19 in Campinas, Brazil. Mathematics, 10, 3644.

Shanker, R. and Shukla K.K., 2017a. Ishita distribution and its applications. Biometrics and Biostatistics International Journal, 5, 39–46.

Shanker, R. and Shukla, K.K., 2017b. Power Akash distribution and its Application. Journal of Applied Quantitative Methods, 12(3), 1-10.

Shukla K.K. and Shanker R., 2018. Power Ishita distribution and its application to model lifetime data. Statistics in Transition New Series, 19, 135–148.

Tahir M.H. and Nadarajah S., 2015. Parameter induction in continuous univariate distributions: Well-established G families. Anais da Academia Brasileira de Ciências, 87, 539–568.

Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica, 57, 307–333.

Wolfram Research, 2020. Mathematica. Version 12.2. Champaign, IL.

Zografos, K. and Balakrishnan, N., 2009. On families of beta- and generalized gamma-generated distributions and associated inference. Statistical Methodology, 6, 344–362.